

Estimation Optimality of Corrected AIC and Modified Cp in Linear Regression

by

Simon L. Davies

Pfizer Global Pharmaceuticals, Inc.

Andrew A. Neath ¹

Department of Mathematics and Statistics, Southern Illinois University Edwardsville

Joseph E. Cavanaugh

Department of Biostatistics, The University of Iowa

Summary

Model selection criteria often arise by constructing unbiased or approximately unbiased estimators of measures known as expected overall discrepancies (Linhart & Zucchini, 1986, p. 19). Such measures quantify the disparity between the true model (i.e., the model which generated the observed data) and a fitted candidate model. For linear regression with normally distributed error terms, the “corrected” Akaike information criterion and the “modified” conceptual predictive statistic have been proposed as exactly unbiased estimators of their respective target discrepancies. We expand on previous work to additionally show that these criteria achieve minimum variance within the class of unbiased estimators.

Key words: AICc, Gauss discrepancy, Kullback-Leibler discrepancy, MC_p , model selection criteria.

¹Corresponding author: Andrew A. Neath, Associate Professor, Department of Mathematics and Statistics, Southern Illinois University, Edwardsville, IL 62026.

1 Introduction

The “corrected” Akaike information criterion, AICc, is an adjusted version of AIC (Akaike, 1973, 1974) originally proposed for linear regression with normal errors (Sugiura, 1978; Hurvich & Tsai, 1989). For fitted candidate models which are correctly specified or overfit, AIC is only asymptotically unbiased while AICc is exactly unbiased as an estimator of their target measure, the expected overall Kullback-Leibler discrepancy.

Similarly, the “modified” conceptual predictive statistic, MC_p , is an alternate version of C_p (Mallows, 1973) appropriate for normal linear regression (Fujikoshi & Satoh, 1997). The target measure for MC_p and C_p is the expected overall Gauss discrepancy. C_p is asymptotically unbiased as an estimator of this measure for fitted candidate models which are correctly specified or overfit. MC_p is exactly unbiased regardless of whether the fitted candidate model is underfit, correctly specified, or overfit, but this unbiasedness requires that the true model is contained in the candidate family.

In this paper, we expand on previous work that has established the exact unbiasedness of AICc (Sugiura, 1978; Hurvich & Tsai, 1989) and MC_p (Fujikoshi & Satoh, 1997) in the context of normal linear regression. We additionally show that these criteria achieve minimum variance within the class of unbiased estimators. Thus, we establish that AICc and MC_p are the minimum variance unbiased estimators (MVUEs) of their respective target discrepancies.

The next section provides an outline of model selection in the linear regression setting. In Section 3, we prove that AICc is the MVUE of the expected overall Kullback-Leibler discrepancy. In Section 4, we show that MC_p is the MVUE of the expected overall Gauss discrepancy. The justifications are based on the traditional approach of appealing to the Lehmann-Scheffé theorem. Section 5 concludes, and features a discussion of some of the practical implications of our results.

2 Outline of Model Selection Framework

Consider a collection of data y , generated according to the normal linear model

$$y = X_o \beta_o + e_o, \quad e_o \sim N_n(0, \sigma_o^2 I). \quad (2.1)$$

We assume that the response vector y is $(n \times 1)$, the design matrix X_o is $(n \times p_o)$ of full column rank, and the parameter vector β_o is $(p_o \times 1)$. Let $C(X_o)$ denote the column space of X_o . Let $\theta_o = (\beta_o', \sigma_o^2)'$ represent the $(p_o + 1)$ -dimensional parameter vector.

Our goal is to search among a class of postulated models for the fitted model which serves as the “best” approximation to model (2.1). Suppose each postulated model is of the form

$$y = X\beta + e, \quad e \sim N_n(0, \sigma^2 I). \quad (2.2)$$

Here, we assume that X is $(n \times p)$ of full column rank, and β is $(p \times 1)$. Denote the corresponding column space $C(X)$ and projection matrix $H = X(X'X)^{-1}X'$. Let $L(\theta | y)$ represent the likelihood function for model (2.2), where $\theta = (\beta', \sigma^2)'$ is the $(p + 1)$ -dimensional parameter vector. Let $\hat{\theta} = (\hat{\beta}', \hat{\sigma}^2)'$ denote the vector of estimates obtained by maximizing $L(\theta | y)$ over the parameter space Θ .

We refer to (2.1) as the *true model* and to (2.2) as a *candidate model*. Let E_o denote the expectation under the true model (2.1).

The collection of all candidate models of interest is known as the *candidate family*. This family will generally consist of models with design matrices of differing ranks based on various sets of regressor variables.

To determine which of the fitted models in the candidate family best resembles the true model, we require a measure which provides a suitable reflection of the disparity between the true model and a fitted candidate model. The overall Kullback-Leibler discrepancy and the overall Gauss discrepancy both fulfill this objective. These measures are introduced in the following sections.

3 The Kullback-Leibler Discrepancy and AICc

The overall Kullback-Leibler discrepancy is constructed using the Kullback-Leibler information (Kullback, 1968) between the true model (2.1) and the candidate model (2.2). The measure is defined as

$$d_{\text{KL}}(\hat{\theta}, \theta_o) = E_o\{-2 \ln L(\theta | y)\}_{|\theta=\hat{\theta}}.$$

(See Linhart & Zucchini, 1986, pp. 11, 18.) The expected overall Kullback-Leibler discrepancy is therefore defined as

$$\begin{aligned} \Delta_{\text{KL}}(\theta_o, p) &= E_o\{d_{\text{KL}}(\hat{\theta}, \theta_o)\} \\ &= E_o\{E_o\{-2 \ln L(\theta | y)\}_{|\theta=\hat{\theta}}\}. \end{aligned}$$

Model selection criteria based on $d_{\text{KL}}(\hat{\theta}, \theta_o)$ are developed by finding a statistic that has an expectation which is equal to, or at least approximately equal to, the expected overall Kullback-Leibler discrepancy. Thus, the targeted measure is $\Delta_{\text{KL}}(\theta_o, p)$.

In the linear regression setting with normal errors, “corrected” AIC is defined as

$$\text{AICc} = n \ln \hat{\sigma}^2 + \frac{n(n+p)}{n-p-2}, \quad (3.1)$$

where $\hat{\sigma}^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/n$ and $\hat{\beta} = (X'X)^{-1}X'y$.

Under the assumption that $C(X_o) \subseteq C(X)$, AICc serves as an exactly unbiased estimator for $\Delta_{\text{KL}}(\theta_o, p)$. (The condition $C(X_o) \subseteq C(X)$ is equivalent to assuming that the candidate model (2.2) subsumes the true model (2.1), or to requiring that $\theta_o \in \Theta$.) The unbiasedness property is justified in the development which follows.

Neglecting the additive constant involving 2π , the log-likelihood for the candidate model (2.2) is given by

$$\ln L(\theta | y) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta).$$

Under the true model (2.1), we have the following expectation:

$$\begin{aligned} E_o\{-2 \ln L(\theta | y)\} &= n \ln \sigma^2 + \frac{1}{\sigma^2} E_o\{(y - X\beta)'(y - X\beta)\} \\ &= n \ln \sigma^2 + \frac{n\sigma_o^2}{\sigma^2} + \frac{1}{\sigma^2} (X_o\beta_o - X\beta)'(X_o\beta_o - X\beta). \end{aligned} \quad (3.2)$$

Evaluating (3.2) at $\theta = \hat{\theta}$ yields

$$E_o\{-2 \ln L(\theta | y)\}_{\theta=\hat{\theta}} = n \ln \hat{\sigma}^2 + \frac{n\sigma_o^2}{\hat{\sigma}^2} + \frac{1}{\hat{\sigma}^2}(X_o\beta_o - X\hat{\beta})'(X_o\beta_o - X\hat{\beta}). \quad (3.3)$$

We can write $X\hat{\beta} = Hy$ and $\hat{\sigma}^2 = y'(I - H)y/n$. An expression for $d_{\text{KL}}(\hat{\theta}, \theta_o)$ is provided by (3.3).

To evaluate $\Delta_{\text{KL}}(\theta_o, p)$, we use the property $HX_o = X_o$, which follows from the assumption $C(X_o) \subseteq C(X)$. Note that $(n\hat{\sigma}^2/\sigma_o^2)$ and the quadratic form

$$\frac{1}{\sigma_o^2}(X_o\beta_o - X\hat{\beta})'(X_o\beta_o - X\hat{\beta})$$

both have chi-square distributions, with degrees of freedom given by $(n - p)$ and p , respectively. Using these results along with the independence of $\hat{\sigma}^2$ and $\hat{\beta}$, the expectation of (3.3) under (2.1) reduces to

$$E_o\{E_o\{-2 \ln L(\theta | y)\}_{\theta=\hat{\theta}}\} = E_o\{n \ln \hat{\sigma}^2\} + \frac{n^2}{n - p - 2} + \frac{np}{n - p - 2}. \quad (3.4)$$

With reference to (3.1), the preceding establishes

$$E_o\{\text{AICc}\} = \Delta_{\text{KL}}(\theta_o, p),$$

thereby proving that AICc is exactly unbiased for $\Delta_{\text{KL}}(\theta_o, p)$, the expected overall Kullback-Leibler discrepancy.

The expectation in (3.4) can be further simplified by expressing $E_o\{n \ln \hat{\sigma}^2\}$ solely in terms of n , p , σ_o^2 , and the *digamma* or *psi* function ψ (Kotz *et al.*, 1982, p. 373). Using the fact that

$$E_o\left\{\ln\left(\frac{n\hat{\sigma}^2}{\sigma_o^2}\right)\right\} = \ln 2 + \psi\left(\frac{n - p}{2}\right),$$

we have

$$\Delta_{\text{KL}}(\theta_o, p) = n \ln \sigma_o^2 + n \ln \frac{2}{n} + n\psi\left(\frac{n - p}{2}\right) + \frac{n(n + p)}{n - p - 2}. \quad (3.5)$$

(See Hurvich & Tsai, 1989, p. 303.)

To show AICc is the minimum variance unbiased estimator of the expected overall Kullback-Leibler discrepancy, we will make use of the Lehmann-Scheffé theorem. In the

present setting, the theorem ensures that if there exists an unbiased estimator of $\Delta_{\text{KL}}(\theta_o, p)$ which is a function of a complete sufficient statistic, then it is the MVUE of the discrepancy. We have already established AICc as an unbiased estimator, and so it remains to be shown that AICc is a function of a complete sufficient statistic.

The likelihood function for the candidate model (2.2) is given by

$$L(\theta | y) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} (y - X\beta)' (y - X\beta) \right\}. \quad (3.6)$$

This likelihood can be re-written in the following form:

$$\begin{aligned} L(\theta | y) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (y'y - 2\beta' X'y + \beta' (X'X)\beta) \right\} \\ &= c(\theta) \exp \left\{ -\frac{1}{2\sigma^2} (y'y) + \left(\frac{1}{\sigma^2} \beta' \right) (X'y) \right\}, \end{aligned}$$

where

$$c(\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \beta' (X'X)\beta \right\}.$$

For the development of AICc, we consider a setting where it is only known that the candidate model is not underspecified. In this context, there are no linear constraints on either the data $(y'y, X'y)$ or the parameters

$$\left(-\frac{1}{2\sigma^2}, \frac{1}{\sigma^2} \beta \right).$$

Thus, by Theorem 2.5.3 of Christensen (1996, pp. 26-27), $T(y) = (y'y, X'y)$ is a complete sufficient statistic.

AICc is defined in (3.1). Note that $n\hat{\sigma}^2 = y'(I - H)y = (y'y) - (X'y)'(X'X)^{-1}(X'y)$ is a function of the complete sufficient statistic $T(y)$. Hence, by the Lehmann-Scheffé theorem, we can conclude that AICc is the minimum variance unbiased estimator for $\Delta_{\text{KL}}(\theta_o, p)$.

4 The Gauss Discrepancy and MC_p

The overall Gauss discrepancy reflects the squared distance between the mean response vector under the true model and under a fitted candidate model. In the regression context, the measure is given by

$$(X_o\beta_o - X\hat{\beta})'(X_o\beta_o - X\hat{\beta}).$$

(See Linhart & Zucchini, 1986, pp. 11, 18-19, 118.) For our purposes, we scale the preceding measure by the true error variance σ_o^2 , thereby obtaining

$$d_G(\hat{\theta}, \theta_o) = \frac{1}{\sigma_o^2} (X_o \beta_o - X \hat{\beta})' (X_o \beta_o - X \hat{\beta}).$$

The expected overall Gauss discrepancy is therefore defined as

$$\begin{aligned} \Delta_G(\theta_o, p) &= E_o \{d_G(\hat{\theta}, \theta_o)\} \\ &= E_o \left\{ \frac{1}{\sigma_o^2} (X_o \beta_o - X \hat{\beta})' (X_o \beta_o - X \hat{\beta}) \right\}. \end{aligned} \quad (4.1)$$

Model selection criteria based on $d_G(\hat{\theta}, \theta_o)$ are developed by finding a statistic that has an expectation which is equal to, or at least approximately equal to, the expected overall Gauss discrepancy. Thus, the targeted measure is $\Delta_G(\theta_o, p)$.

Let X_* denote the design matrix for the largest model in the candidate family. Assume X_* is $(n \times p_*)$ of full column rank. Let $\theta_* = (\beta_*', \sigma_*^2)'$ denote the $(p_* + 1)$ -dimensional parameter vector for the candidate model (2.2) associated with X_* , and let Θ_* represent the corresponding parameter space. Let $\hat{\theta}_* = (\hat{\beta}_*', \hat{\sigma}_*^2)'$ denote the maximum likelihood estimator of θ_* .

Let $C(X_*)$ denote the column space of X_* , and let $H_* = X_*(X_*'X_*)^{-1}X_*'$ project onto $C(X_*)$. We will henceforth assume $C(X_o) \subseteq C(X_*)$. (This is equivalent to assuming that the largest model in the candidate family subsumes the true model, or to requiring that $\theta_o \in \Theta_*$.)

In the linear regression setting with normal errors, the “modified” conceptual predictive statistic is defined as

$$MC_p = \frac{(n - p_* - 2)\hat{\sigma}^2}{\hat{\sigma}_*^2} + 2p - n + 2. \quad (4.2)$$

Under the assumption that $C(X_o) \subseteq C(X_*)$, MC_p serves as an exactly unbiased estimator for $\Delta_G(\theta_o, p)$. The unbiasedness property is justified in the following development.

First, we derive an expression for $\Delta_G(\theta_o, p)$. We have

$$(X_o \beta_o - X \hat{\beta})' (X_o \beta_o - X \hat{\beta}) = (y - X_o \beta_o)' H (y - X_o \beta_o) + (X_o \beta_o)' (I - H) (X_o \beta_o). \quad (4.3)$$

Under the true model (2.1), the expected value of the first term in (4.3) is

$$E_o\{(y - X_o\beta_o)'H(y - X_o\beta_o)\} = p\sigma_o^2. \quad (4.4)$$

Using (4.3) and (4.4) with (4.1), we obtain

$$\Delta_G(\theta_o, p) = p + \frac{(X_o\beta_o)'(I - H)(X_o\beta_o)}{\sigma_o^2}. \quad (4.5)$$

Next, we establish that the expected value of MC_p under (2.1) is equal to (4.5). We begin by deriving the expected value of the ratio $\hat{\sigma}^2/\hat{\sigma}_*^2$.

Note that

$$\frac{\hat{\sigma}^2}{\hat{\sigma}_*^2} = 1 + \frac{y'(H_* - H)y}{y'(I - H_*)y}. \quad (4.6)$$

Since $C(X) \subseteq C(X_*)$, $H_*H = HH_* = H$, and thus the quadratic forms $y'(H_* - H)y$ and $y'(I - H_*)y$ are independent.

Since $C(X_o) \subseteq C(X_*)$, we have $H_*X_o = X_o$. Thus, we can establish the following two results:

$$E_o\{y'(H_* - H)y\} = \sigma_o^2(p_* - p) + (X_o\beta_o)'(I - H)(X_o\beta_o), \quad (4.7)$$

$$\frac{y'(I - H_*)y}{\sigma_o^2} \sim \chi^2(n - p_*). \quad (4.8)$$

From (4.8), we have

$$E_o\left\{\frac{\sigma_o^2}{y'(I - H_*)y}\right\} = \frac{1}{n - p_* - 2}. \quad (4.9)$$

Using the independence of $y'(H_* - H)y$ and $y'(I - H_*)y$ along with (4.7) and (4.9), the expectation of (4.6) reduces to

$$E_o\left\{\frac{\hat{\sigma}^2}{\hat{\sigma}_*^2}\right\} = \frac{n - p - 2}{n - p_* - 2} + \frac{(X_o\beta_o)'(I - H)(X_o\beta_o)}{(n - p_* - 2)\sigma_o^2}. \quad (4.10)$$

Then by substituting (4.10) into (4.2), we obtain

$$E_o\{MC_p\} = p + \frac{(X_o\beta_o)'(I - H)(X_o\beta_o)}{\sigma_o^2}.$$

By (4.5), the preceding is equal to $\Delta_G(\theta_o, p)$. Thus, we have

$$E_o\{MC_p\} = \Delta_G(\theta_o, p).$$

We have thereby established that MC_p is exactly unbiased for $\Delta_G(\theta_o, p)$, the expected overall Gauss discrepancy. Note that this unbiasedness property depends only on the largest candidate model and holds irrespective of whether the candidate model of interest is under-specified, correctly specified, or overspecified.

We now look to establish a minimum variance unbiasedness property for MC_p . Once again, we shall make use of the Lehmann-Scheffé theorem.

For the estimation of $\Delta_G(\theta_o, p)$, the likelihood must be formulated differently than the likelihood (3.6) used in the previous section for the estimation of $\Delta_{KL}(\theta_o, p)$. The unbiasedness justification of AICc requires $C(X_o) \subseteq C(X)$. Thus, the candidate model of interest is not underspecified, although any smaller candidate model potentially is. The likelihood is thereby formulated using the candidate model of interest. The estimate of the true error variance σ_o^2 in expression (3.5) for $\Delta_{KL}(\theta_o, p)$ is based on the complete sufficient statistic corresponding to this likelihood.

The unbiasedness justification of MC_p is more general, requiring $C(X_o) \subseteq C(X_*)$ but not $C(X_o) \subseteq C(X)$. Thus, the largest model in the candidate family is not underspecified, although any smaller candidate model (including the model of interest) potentially is. The likelihood is thereby formulated using the largest candidate model. The estimates of the true regression parameter vector β_o and the true error variance σ_o^2 in expression (4.5) for $\Delta_G(\theta_o, p)$ are based on the complete sufficient statistic corresponding to this likelihood.

This complete sufficient statistic is easily seen to be $T_*(y) = (y'y, X_*'y)$. Once again, we apply Theorem 2.5.3 of Christensen (1996) to establish this fact. In the development of MC_p , we consider a setting where it is only known that the largest candidate model is not underspecified. In this context, there are no linear constraints on either the data or the parameters that preclude the application of this theorem.

The residual sum of squares for the largest model is calculated as

$$SS_{Res}^* = y'(I - H_*)y = y'y - (H_*y)'(H_*y). \quad (4.11)$$

Since $H_*y = X_*(X_*'X_*)^{-1}X_*'y$ is a function of $X_*'y$, we can see from (4.11) that SS_{Res}^* is a

function of $T_*(y)$.

Noting $C(X) \subseteq C(X_*)$ so that $H_*H = HH_* = H$, we can express the residual sum of squares for the candidate model as

$$SS_{Res} = y'(I - H)y = y'y - (Hy)'(Hy) = y'y - (HH_*y)'(HH_*y). \quad (4.12)$$

Since HH_*y is a function of X'_*y , SS_{Res} is a function of $T_*(y)$. From (4.11), (4.12), and (4.2), it thereby follows that $(\hat{\sigma}^2/\hat{\sigma}_*^2) = (SS_{Res}/SS_{Res}^*)$ and MC_p are functions of $T_*(y)$.

Hence, MC_p is an exactly unbiased estimator of $\Delta_G(\theta_o, p)$ and is a function of the complete sufficient statistic. By the Lehmann-Scheffé theorem, we can conclude MC_p is the minimum variance unbiased estimator for $\Delta_G(\theta_o, p)$.

5 Concluding Remarks

An important topic in regression theory is the statistical problem of determining which input variables are needed for estimating a response function. Such a decision is often facilitated by the use of a model selection criterion. One approach to developing a selection criterion is to estimate an expected overall discrepancy, a quantity that reflects the degree of similarity between the true model and a fitted model. The choice is then the fitted model computed to be most similar to the unknown truth.

Two well-studied choices for a discrepancy measure are the Kullback-Leibler discrepancy and the Gauss discrepancy. For the corresponding expected overall discrepancies, unbiased estimators have been proposed via the statistics AICc and MC_p . A subsequent question arises as to whether these model selection criteria attain a type of optimality in terms of variability. We have shown that AICc and MC_p are indeed the best choices for unbiased estimation of their respective target discrepancies.

The conditions under which AICc and MC_p are minimum variance unbiased estimators differ in the formulation of the underlying likelihood function. For AICc, we consider a setting where it is only known that the candidate model is not underspecified. In this setting, the expected overall Kullback-Leibler discrepancy $\Delta_{KL}(\theta_o, p)$ is defined through $n \ln \sigma_o^2$, the

natural log of the true error variance. AICc is developed using an estimator of this parameter under conditions set forth by the candidate model likelihood function. For MC_p , we consider a setting where it is only known that the largest candidate model is not underspecified. In this setting, the expected overall Gauss discrepancy $\Delta_G(\theta_o, p)$ is defined through $(X_o\beta_o)'(I - H)(X_o\beta_o)/\sigma_o^2$, a parameter whose value depends on the candidate model of interest via the projection matrix H . MC_p is developed using an estimator of this parameter based on the particular candidate model under conditions set forth by the largest candidate model. Roughly speaking, values of AICc arise from estimating a common parameter under different models in the candidate family, whereas values of MC_p arise from estimating a class of parameters under the same model (i.e., the largest candidate model).

In practice, model selection is accomplished by choosing that model from the candidate family for which the criterion, either AICc or MC_p , is minimized. It should be noted that although AICc and MC_p are unbiased with respect to a specific candidate model, the minimum of the criterion over the candidate family is not unbiased for the expected overall discrepancy corresponding to the selected model. The bias will clearly grow as the probability of selecting an incorrect candidate model increases. In general, the probability of choosing an inappropriate model increases in accordance with the size of the candidate family. This consideration emphasizes the importance of a careful and deliberate *a priori* formulation of the candidate family, where every model is based on a sensible combination of plausible explanatory variables.

In establishing the minimum variance unbiasedness property for AICc, we assume it is only known that the candidate model of interest is not underspecified. In a practical application when AICc is computed for a particular candidate model, one cannot determine whether this condition is satisfied. However, for models where the condition is clearly violated, the optimality of AICc is less important than for models where the condition might apply. Models that are grossly underspecified will correspond to large values of the expected Kullback-Leibler discrepancy, thereby resulting in large values of AICc. Such models may be easily distinguished from models that exhibit fidelity to the data. On the other hand,

models providing adequate fit may be difficult to distinguish among. Precise estimation of the Kullback-Leibler discrepancy is therefore more imperative for models where the degree of underspecification is marginal or nonexistent. For any such model, the assumption under which we have established the optimality of AICc is defensible.

In demonstrating the minimum variance unbiasedness property for MC_p , we assume it is only known that the largest candidate model is not underspecified. Again, in a practical application, one cannot determine whether this condition is satisfied. However, with MC_p , the optimality of the criterion hinges on the propriety of the largest candidate model. Thus, this model must be based on a judicious choice of regressor variables. In particular, if the largest candidate model is quite possibly underspecified due to the exclusion of potentially relevant regressor variables, then the preceding condition is not defensible. On the other hand, as previously mentioned, it is clearly not advisable to include regressor variables of questionable merit in the formulation of candidate models merely to ensure that the largest candidate model is not underspecified.

Acknowledgment

The authors would like to express their appreciation to the referees for providing thoughtful and insightful comments which helped to improve the original version of this manuscript.

References

- Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In *2nd International Symposium on Information Theory*, Eds. B. N. Petrov and F. Csáki, pp. 267–281. Budapest: Akadémia Kiadó.
- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, **AC-19**, 716–723.
- Christensen, R. (1996). *Plane Answers to Complex Questions (2nd Edition)*. New York: Springer.
- Fujikoshi, Y., and Satoh, K. (1997). Modified AIC and C_p in Multivariate Linear Regression. *Biometrika*, **84**, 707–716.
- Hurvich, C. M. and Tsai, C. L. (1989). Regression and Time Series Model Selection in Small Samples. *Biometrika*, **76**, 297–307.
- Kotz, S., Johnson, N. L., and Read, C. B., Eds. (1982). *Encyclopedia of Statistical Sciences, Volume 2*. New York: Wiley.
- Kullback, S. (1968). *Information Theory and Statistics*. New York: Dover.
- Linhart, H. and Zucchini, W. (1986). *Model Selection*. New York: Wiley.
- Mallows, C.L. (1973). Some Comments on C_p . *Technometrics*, **15**, 661–675.
- Sugiura, N. (1978). Further Analysis of the Data by Akaike’s Information Criterion and the Finite Corrections. *Communications in Statistics*, **A7**, 13–26.

Résumé

Les critères de modèle de sélection naissent souvent de la construction de mesures d'estimation impartiales, ou approximativement impartiales, connues comme divergences globales prévues. De telles mesures quantifient la disparité entre le vrai modèle (c'est-à-dire le modèle qui a produit les données observées) et un modèle candidat correspondant. En ce qui concerne les applications de régression linéaires contenant des erreurs distribuées normalement, le modèle de critère d'information "corrigé" Akaike et le modèle conceptuel de statistique de prévision "modifié" ont été proposés comme étant des instruments exacts de mesures d'estimation impartiales de leurs objectifs respectifs de divergences. En nous appuyant sur les travaux précédents et en les développant, nous proposons de démontrer, en outre, que ces critères réalisent une variance minimum au sein de la classe des instruments de mesures d'estimation impartiales.